STUDY
NOTE
2004-04

Fairness of Army ASVAB Test Composites for MOS and Job Families

Joseph Zeidner, Cecil Johnson,
Yefim Vladimirsky, and Susan Weldon
J. Zeidner & Associates

United States Army Research Institute
for the Behavioral and Social Sciences

April 2004

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the U.S. Army Human Resources Command

ZITA M. SIMUTIS
Director

Research accomplished under contract
for the Department of the Army

J. Zeidner & Associates

Technical Review by

Peter Greenston

NOTICES

DISTRIBUTION: Primary distribution of this Study Note has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-PO, 5001 Eisenhower Ave., Alexandria, VA 22304-4841.

FINAL DISPOSITION: This Study Note may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Study Note are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) April 2004	2. REPORT TYPE Interim	3. DATES COVERED (from. . . to) April 2002 – August 2003
4. TITLE AND SUBTITLE Fairness of Army ASVAB Test Composites for MOS and Job Families		5a. CONTRACT OR GRANT NUMBER DASW01-02-P-0355
		5b. PROGRAM ELEMENT NUMBER 65803
6. AUTHOR(S) Joseph Zeidner, Cecil Johnson, Yefim Vladimirsky, Susan Weldon		5c. PROJECT NUMBER D730
		5d. TASK NUMBER 263
		5e. WORK UNIT NUMBER C01
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) J. Zeidner & Associates 5621 Old Chester Ct. Bethesda, MD 20814		8. PERFORMING ORGANIZATION REPORT NUMBER
18. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22304-4841		18. MONITOR ACRONYM ARI
		11. MONITOR REPORT NUMBER

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words*

Major goals of this research are to evaluate the fairness to minorities of the new operational ASVAB seven test composites and to compare the seven test composites with the previously utilized nine test composites.

The experiment evaluates fairness in predicting performance outcomes as measured by the actual SQT (Skill Qualification Test) scores of individuals assigned to MOS by the operational system. Fairness is measured by the difference between predicted performance and SQT scores across gender and race. Complete fairness is indicated by very small mean difference and fairness to minorities is present when mean difference in the minority group is zero or has a positive sign (overprediction). The sample uses ARI's data set of ASVAB scores limited to 66 MOS and SQT scores obtained during FY 1987-1989. The total sample size of first-term enlistees was 83,132.

Findings indicate that the proposed new seven least squares estimate (LSE) test composites are comparatively fair to minorities. The results using seven test composites are comparable to the previous study using nine test composites.

15. SUBJECT TERMS
Armed Services Vocational Aptitude Battery (ASVAB); Army aptitude area composites; personnel classification; fairness to minorities; Differential Assignment Theory

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unclassified	40	Dr. Peter Greenston, 703.617.0344

ACKNOWLEDGEMENT

The authors thank Dr. Peter Greenston of the Army Research Institute for his contributions to the research effort, including his support, encouragement, and careful review of the draft report.

Fairness of New Army ASVAB Test Composites for MOS and Job Families

EXECUTIVE SUMMARY

Requirement

A major objective of the present research is to determine whether the current seven test Armed Services Vocational Aptitude Battery (ASVAB) composites have at least as much fairness for black and female recruits as did the nine test ASVAB composites which they replaced. Fairness is traditionally defined as the absence of underpredictions for the minority groups which are considered potentially susceptible to discrimination.

Procedures

The experiment used data obtained from the operational system in terms of its actual assignment of recruits to military occupational specialties (MOS) and second tier job families. Two different test composites, one based on seven and the other on nine tests, were evaluated to determine their fairness in predicting performance outcomes as measured by the actual SQT scores of each individual assigned to one of 66 MOS. Fairness is measured by the difference between predicted performance and SQT scores (the prediction error scores) for males and females and for whites and blacks. Complete fairness is indicated by very small or zero mean differences. Fairness with respect to minorities is indicated when the differences approach zero or indicate some (small) overprediction.

Findings

In evaluating test composite prediction error scores (PEs) resulting from operational assignment to MOS and job families, a distinct trend of underprediction was found for blacks and females, but the level of underprediction was too small to have practical significance. In

testing for statistical significance, PEs for 12 MOS were found to have statistically significant ($p=.05$) critical ratio differences from zero for blacks. The previous study also found several significant differences for blacks. For females, statistically significant underpredictions were found for 17 MOS. Again, this was found in the previous study that had fewer MOS. More importantly, in testing the mean difference, the prediction error differences for blacks and females were minor. For blacks, the overall mean prediction error was found to be –0.019, or .38 in Army standard score units (mean of 100 and a S.D. of 20). For females, the mean prediction error was found to be –0.108 or 2.16 Army standard score units.

The findings for minorities are consistent with research findings in the civilian employment and military settings concerning regression-line differences in intercept values. Such differences, in the same direction as found in this study, appear to be a relatively common phenomenon. In several instances of comparisons between groups, the use of regression equations computed on the total group gives advantage to minority groups members, i.e., minority groups are overpredicted. In the present study, however, minority groups were underpredicted even though predictions were based on the total group, as is appropriate for an operational system.

The overall conclusion, then, is that the proposed new seven test LSE composites are fair to minorities while providing substantial improvements in classification efficiency.

Fairness of New Army ASVAB Test Composites for MOS and Job Families

Table of Contents

Tables and Figures:

INTRODUCTION

Background

The objective of a previous research effort (Zeidner, Johnson, Vladimirsky and Weldon, 1998) was undertaken to determine the effect of proposed least-squares-estimate (LSE) ASVAB test composites on gender and racial fairness.[1] The previous study differed from most previous studies in a number of respects including: (1) replacement of a selection with a classification paradigm; (2) replacement of *g* loaded predictors with LSEs specific to each job and job family for use in estimating predicted performance; (3) examination of the joint consequences of fairness; (4) use of prediction error scores in precisely measuring the degree of over- or underprediction rather than just observing the occurrence of these phenomena; and (5) use of 66 MOS best weighted (tailored) test composites in a classification design rather than the use of a small number of test composites in a selection design. The research utilized the nine tests (also referred to as subtests) that constituted the ASVAB at that time.

Civilian and Military Settings

As noted, fairness is traditionally defined as the absence of underprediction for the minority group for which discrimination potentially exists (Cleary, 1986). Thus, if a test is used for selection and is underpredicting minority groups, members of a minority group may be rejected when they are capable of adequate performance. Thorndike proposed a modified model that stipulates a selection measure as being fair only if the success ratio equals the selection ratio for minorities (1971).

[1] Another major objective of the previous study was the development of a new methodology for making decisions based on trade-offs between classification efficiency and fairness to minorities in the design and use of a new system that optimally assigns first-term recruits to jobs.

A major objective of the previous research (Zeidner, et al., 1998) was to provide evidence that the proposed nine test LSE composites have at least as much fairness for black and female recruits as did the unit weighted aptitude area composites (in effect prior to January 2002). The LSE composites in the previous study employed nine tests as compared to seven tests utilized in defining the composites in the present study.

In the military context, Wise, Welsh, Grafton, Foley, et al. (1992) found that whites had significantly higher expected criterion scores than blacks. The authors stated that while the differences are of statistical significance (in these large samples), they are of limited practical significance, being only about one-tenth of a standard deviation. Most of the individual composites also showed significant overpredictions for blacks (p. 23). The overall results also showed that males had higher criterion scores than females (except at the highest level of the composite scores for the job family). Two Army composites, general maintenance (GM) and surveillance and communication (SC), showed significant underpredictions for blacks. But here again, the authors note that the size of the differences were quite small despite statistical significance.

Finding of the Previous Study

Findings of the previous study showed that there was a distinct trend of negative prediction error (underpredictions) for females and blacks, although these differences were so small as to be considered of little practical importance. The overall mean prediction correction was comparable to 0.5 points in Army Aptitude Area standard score units for blacks and 1.7 points for females. Aptitude Areas have a mean of 100 and a standard deviation of 20.

Findings were consistent with results found in the civilian employment sector and in the military only with regard to differences in regression-line intercept values between minority groups and the majority. When predictions of minority performance are based on the total group in these sectors, minority groups are generally overpredicted. In our previous study, however, minority groups were underpredicted even though predictions were based on the total group (as is appropriate for the operational system). In the previous study, the percentage of minorities in the total group was much greater than in most other studies available for comparison, thus producing smaller changes in intercept values for minority groups. Comparisons of the existing unit-weighted aptitude area composites with the proposed new LSE test composites resulted in much smaller prediction errors for the latter composites in all groups.

Since the overall finding was that the proposed LSE nine test composites would improve fairness to minorities while improving classification efficiency, these issues should not impede consideration of a new classification system that utilizes LSEs to optimally assign recruits to all Army entry-level jobs.

General Goal

The Army adopted "interim" Aptitude Area (AA) composites in January, 2002. The interim composites reflect a nine job family structure based on composites of seven ASVAB tests, where least squares regression is used to estimate the weights (applied to ASVAB tests to form the composites). The composites, in turn, are referred to as LSE composites.

Since the publication of the Zeidner, Johnson, Vladimirsky, and Weldon (August, 2000) "Specification" study, DOD decided to remove the Numerical Operations (NO) and Coding Speed (CS) tests from the ASVAB battery. The tests were dropped from the battery in part because of the difficulty of maintaining computer-administered speeded tests and in part because

of the small contribution that NO and CS made to predictive validity in the selection process[2]. It remained to be seen that issues of gender and racial fairness not be impeded by the use of seven test composites. ARI proposed that before additional consideration be given to the "interim" ASVAB that the fairness of the newly proposed seven test composites should be examined. Additional studies within this contract will compare the methods of Cleary and Thorndike with the authors prediction error method.

OBJECTIVES

In the previous study, gender and race performance prediction errors were computed for 66 MOS and 9 job families but were accomplished only for the nine-test LSE composites. The present study was designed to obtain prediction error scores for 66 MOS and nine job families using LSE composites based on seven tests. A double cross-validation design permitting complete unbiased estimates of prediction errors (PEs) was used. The 66 MOS were selected as being the only robust data set extant containing race and gender information for individuals along with other variables available to the researchers.

The specific research objectives were:

1. To compute PEs for the 9-job family and 66 MOS LSE weighted ASVAB composites of seven tests, by gender and race.

2. To compare results, by gender and race, for composites based on seven and nine ASVAB tests.

---

[2] Zeidner, Johnson, Vladimirsky and Weldon (December, 2000) determined the classification effect of reducing the ASVAB by dropping the two speeded tests NO and CS. The unbiased overall MPP for classification was significantly lowered by 6.3 percent in the total sample for the 150 families and by 8.7 percent for the 66 family structure. More recently, the authors conducted a comprehensive examination of validities embracing the proposed first tier (150 job families), the second tier (17 job families) and the interim LSE battery (composed of the seven test LSE test composites for the ASVAB nine operational job families). Composite validities are often used as a conventional index of merit in selection programs and they are also used in the process of establishing cut scores for jobs, generally employing youth validities (Zeidner, Johnson, Vladimirsky, and Weldon, November 2002).

METHOD

General Approach

This study is comprised of an experiment that is similar to typical studies of selection fairness employing the ASVAB. In this experiment, the actual assignments of recruits to jobs made by the operational system are used. Actual Skill Qualification Test (SQT) scores achieved by recruits in their assigned MOS are used as the criterion measure. Assignment variables (AVs), based on the least squares estimates (LSEs) of SQT scores using ASVAB tests, constitute the predicted performance measures. Two different sets of AV composites, one using 7 and the other 9 tests are compared. Fairness for each condition is measured by the difference between predicted performance (AV) and SQT scores for each MOS or job family by gender and race. Figure 1 indicates the typical triple cross design employed to remove sampling error from mean predicted performance (MPP) estimates. However, as noted, in this study we used a double cross design to remove sampling error and obtain unbiased estimates of PEs.

The independent estimate of criterion scores is based on pure least squares estimates (LSEs) using either positive or negative weights. These weights are corrected for restriction in range to the youth population. The estimate of predicted criterion scores uses weights computed in the analysis sample to obtain pure LSEs of the criterion. These weights are permitted to be either positive or negative, in contrast to the LSEs from an independent analysis sample that are used to make assignments. The optimal weights for assignment LSEs (AVs) are all constrained to be non-negative.

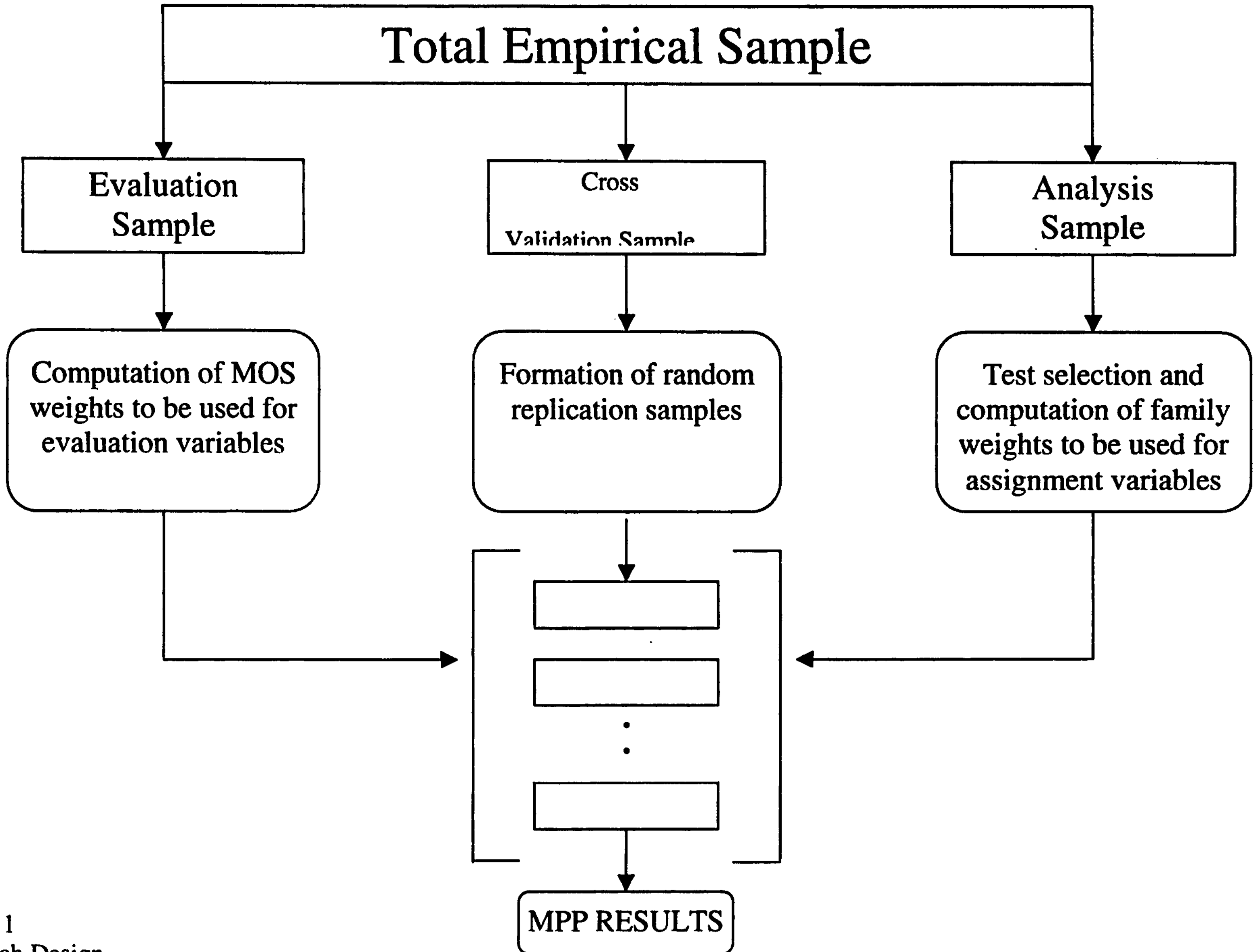

Figure 1
Research Design

SQT Data

In selecting MOS for study, an attempt was made to include as broad a range of entry-level jobs as possible. ARI's data base for SQT years FY 1987-1989 was used to obtain a sample of 83,132 recruits in 66 MOS. The sample in both experiments uses ARI's data set of ASVAB validities and task-based written test, skill level - 1 SQTs.

In 1989 there were about 250 MOS in the Army. Each of these MOS was comprised of one to five skill levels. Prior to 1983, the SQT had both written and hands-on components measuring job proficiency. After 1983 the SQT was designed only as a task-based written test of job proficiency. Soldiers were required to take the SQT annually after completing 11 months or more of service. At the time of this study, FY 1987-1989 data were considered by ARI to be psychometrically good SQT years in terms of discriminability and reliability of the measures.

In a related study by the authors, SQTs were found to be equivalent to specially developed hands-on performance measures used as criteria in Project A (Zeidner, Scholarios, and Johnson; 2003). Equivalence was defined as making the same decisions or as having similar outcomes employing either criterion.

The set of SQT scores in each of the 66 MOS were standardized to have a mean of zero and a standard deviation of one within each MOS.

Correction of Validities

The validities in this study were corrected for restriction in range separately by MOS. Range restriction was due to operational assignment effects, the restriction in range impact of assignment to MOS from a common entry pool. Since this study uses the Army input sample rather than the youth population as the basis for making this correction, no further correction is

made for restriction due to selection effects. Validities are also corrected for unreliability of the criterion variable prior to the restriction in range correction.

Experimental Conditions

Two types of AVs are evaluated separately for 66 and 9 experimental job families. The fairness of the nine and seven test composites are compared for the 66 MOS and the nine experimental job families. As noted, the 66 MOS are the only data that have gender and race associated with each individual.

The score vectors of recruits in the present study are weighted to obtain AVs, with the weights being different for each MOS or job family. In computing standardized AV scores, the data for the total sample in each MOS is used. Thus, the weights to be applied to the score vectors of members of each subsample to form LSE scores is based on the total sample regression weights for that MOS.

Index of Fairness

As noted earlier, the present study uses the prediction error as the index of fairness. The three major indices of fairness proposed by Cleary (1968) use: (1) the slope and (2) the intercept of regression lines representing the prediction of actual performance by measures proposed or already used in the selection process, and (3) the unsquared error of prediction. Some investigators claim that the standard error of estimate also should be considered when regression lines are being compared across groups. The assignment measures used in this study, as a means of directly measuring the unsquared prediction error without computing regression lines, incorporate the effects of regression line slopes and intercepts in making comparisons of fairness across groups. The use of the average PE (MPE) as the index of fairness permits consideration of the entire range of AV scores relevant to classification and initial assignment to jobs. By

avoiding the use of regression lines in our analyses, we also avoid any requirement for considering the effects of disparate standard errors of estimate in making comparisons across groups.

The expected standardized mean prediction error in the total sample (across all MOS and within each MOS) is equal to zero in this study, since these scores are equal to a signed difference obtained by subtracting the standardized SQT score from the standardized AV score of each individual. Both the AV and SQT scores are standardized to have a mean of zero and a SD of one within each MOS for the computation of prediction errors. Prediction error scores, separately by group, i.e., male and female, black and white, are positive in the case of overpredictions for a given condition and negative in the case of underpredictions.

Critical ratios are computed to test the significance of the difference of prediction errors from zero by gender and race separately for the 66 MOS LSE composites. Chi-squares statistical tests of significance were determined for the combined set of MOS (making use of the additivity property of chi squares).

RESULTS AND DISCUSSION

Description of MOS, Job Families and ASVAB

In Table 1 we present descriptive information about the (1989 vintage) 66 MOS used in this study. As noted, these are 66 MOS having gender and race data and, for the most part, relatively large Ns for obtaining stable cross-validities. Assignment variables (AVs) are based on the LSEs of SQT scores using the set of ASVAB tests. There is a different weighted LSE (tailored) test composite for each MOS. Table 1 also shows the names of the 66 MOS used in this study as core jobs forming 66 job families and the percentages of individuals by gender and

race in each MOS. Table 2 shows the 66 MOS rearranged into the same nine experimental job families used in the previous fairness study. The total N for the study was 83,132.

Table 1

Number of First-Term Enlistees Assigned to Each of 66 MOS by Gender and Race in the FY 1987 - 1989 Data Set

MOS	Name	Percent				N
		Male	Female	White	Black	
11B	Infantryman	100.00	0.00	83.27	16.73	3,490
11C	Indirect Fire Infantryman	100.00	0.00	84.65	15.35	1,896
11H	Heavy Anti-Armor Weapons Infantryman	100.00	0.00	85.00	15.00	1,027
11M	Fighting Vehicle Infantryman	100.00	0.00	81.71	18.29	1,416
12C	Bridge Crewmember	100.00	0.00	82.09	17.91	726
13B	Cannon Crewmember	100.00	0.00	54.03	45.97	7,851
13F	Fire Support Specialist	100.00	0.00	81.45	18.55	1,757
13M	Multiple Launch Rocket Sys (MLRS) Crewmember	100.00	0.00	91.20	8.80	375
13N	Lance Crewmember	97.68	2.32	81.65	18.35	474
13R	FA Firefinder Radar Operator	100.00	0.00	82.72	17.28	162
16D	Hawk Missile Crewmember	88.53	11.47	82.44	17.56	279
16P	Chaparral Crewmember	100.00	0.00	84.00	16.00	450
16R	Vulcan Crewmember	100.00	0.00	85.21	14.79	399
16S	Man Portable Air Defense System Crewmember	100.00	0.00	62.25	37.75	837
19E	M48 - M60 Armor Crewman	100.00	0.00	78.15	21.85	1,661
19K	M1 Abrams Armor Crewman	100.00	0.00	81.87	18.13	2,714
29E	Radio Repairer	93.16	6.84	83.80	16.20	395
29J	Telecommunications Terminal Device Repairer	94.87	5.13	86.08	13.92	273
29N	Telephone Central Office Repairer	91.53	8.47	64.50	35.50	307
29V	Strategic Microwave Systems Repairer	90.60	9.40	89.26	10.74	149
31C	Single Channel Radio Operator	91.12	8.88	83.69	16.31	2,839
31K	Combat Signaler	92.04	7.96	60.62	39.38	2,750
31L	Wire Systems Installer	78.84	21.16	52.44	47.56	1,087
31V	Unit Level Communications Maintainer	92.48	7.52	70.79	29.21	1,729

Number of First-Term Enlistees Assigned to Each of 66 MOS by Gender and Race in the FY 1987 - 1989 Data Set

MOS	Name	Percent Male	Percent Female	Percent White	Percent Black	N
33T	EW/I Tactical Systems Repairer	95.77	4.23	98.59	1.41	71
35K	Avionic Mechanic	81.73	18.27	80.71	19.29	197
43E	Parachute Rigger	93.90	6.10	84.08	15.92	377
44B	Metal Worker.	98.32	1.68	80.34	19.66	417
44E	Machinist	99.15	0.85	92.74	7.26	234
45K	Tank Turret Repairer	97.87	2.13	83.54	16.46	328
51B	Carpentry and Masonry Specialist	97.67	2.33	81.14	18.86	859
52D	Power Generator Equipment Repairer	95.45	4.55	78.20	21.80	2,394
54B	Chemical Operations Specialist	92.30	7.70	72.45	27.55	1,078
55B	Ammunitions Specialist	91.40	8.60	72.25	27.75	919
62B	Construction Equipment Repairer	97.06	2.94	76.94	23.06	1,123
62E	Heavy Construction Equipment Operator	98.98	1.02	87.70	12.30	683
62J	General Construction Equipment Operator	98.95	1.05	81.41	18.59	382
63B	Light-Wheel Vehicle Mechanic	91.01	8.99	73.98	26.02	4,439
63E	M1 Abrams Tank System Mechanic	100.00	0.00	86.48	13.52	540
63G	Fuel and Electrical system Repairer	96.46	3.54	83.92	16.08	311
63S	Heavy-Wheel Vehicle Mechanic	98.31	1.69	92.19	7.81	947
63T	Bradley Fighting Vehicle Systems Mechanic	100.00	0.00	94.57	5.43	700
67V	Observation/Scout Helicopter Repairer	97.89	2.11	91.94	8.06	757
68B	Aircraft Powerplant Repairer	83.98	16.02	91.80	8.20	256
68G	Aircraft Structural Repairer	98.44	1.56	91.93	8.07	384
68J	Aircraft Armament/Missile Systems Repairer	96.73	3.27	83.65	16.35	367
71D	Legal Specialist	68.73	31.27	80.00	20.00	550
71L	Administrative Specialist	31.11	68.89	46.80	53.20	765
71M	Chaplain Assistant	66.05	33.95	79.05	20.95	377

Number of First-Term Enlistees Assigned to Each of 66 MOS by Gender and Race in the FY 1987 - 1989 Data Set

		Percent				
MOS	Name	Male	Female	White	Black	N
72E	Tactical Telecommunications Ctr Op	78.68	21.32	58.31	41.69	638
72G	Automatic Data Telecommunications Ctr Op	49.92	50.08	59.78	40.22	649
73C	Finance Specialist	56.20	43.80	54.69	45.31	799
74D	Computer/Machine Operator	61.16	38.84	72.17	27.83	327
75B	Personnel Administration Specialist	68.16	31.84	53.24	46.76	1,542
75D	Personnel Records Specialists	34.07	65.93	41.86	58.14	989
76C	Equipment Records and Parts Specialist	94.17	5.83	58.34	41.66	2,403
76Y	Unit Supply Specialist	83.92	16.08	56.84	43.16	4,279
77F	Petroleum Supply Specialist	86.30	13.70	56.96	43.04	2,846
81E	Graphics Documentation Specialist	62.79	37.21	83.72	16.28	129
84B	Still documentation Specialist	88.42	11.58	88.42	11.58	95
84F	Visual Info/Audio Documentation Specialist	63.74	36.26	65.93	34.07	91
88H	Cargo Specialist	87.99	12.01	59.29	40.71	533
88M	Motor Transport Operator	88.64	11.36	68.82	31.18	5,368
91A	Medical Specialist	83.41	16.59	72.96	27.04	1,790
94B	Food Service Specialist	81.04	18.96	46.79	53.21	3,787
95B	Military Police	86.91	13.09	92.11	7.89	2,369
Total		90.27	9.73	70.29	29.71	83,132

Table 2

Nine Experimental Job Families (Based on 66 Core Jobs - Only Core Jobs Shown)	
Administration (cluster 1)	
51B	Carpentry and Masonry Specialist
71D	Legal Specialist
71L	Administrative Specialist
71M	Chaplain Assistant
72G	Automatic Data Telecommunications Ctr Op
73C	Finance Specialist
74D	Computer/Machine Operator
75B	Personnel Administration Specialist
75D	Personnel Records Specialists
76C	Equipment Records and Parts Specialist
76Y	Unit Supply Specialist
81E	Graphics Documentation Specialist
84F	Visual Info/Audio Documentation Specialist
88H	Cargo Specialist
Armor and Combat Specialties (cluster 2)	
12C	Bridge Crewmember
19E	M48 - M60 Armor Crewman
19K	M1 Abrams Armor Crewman
31K	Combat Signaler
72E	Tactical Telecommunications Ctr Op
91A	Medical Specialist
Combat Systems Operations (cluster 3)	
13F	Fire Support Specialist
13M	Multiple Launch Rocket Sys (MLRS) Crewmember
13N	Lance Crewmember
16R	Vulcan Crewmember
16S	Man Portable Air Defense System Crewmember
31C	Single Channel Radio Operator
43E	Parachute Rigger
55B	Ammunitions Specialist
94B	Food Service Specialist

Nine Experimental Job Families (Based on 66 Core Jobs - Only Core Jobs Shown)	
Electronics Repair (cluster 4)	
29E	Radio Repairer
29J	Telecommunications Terminal Device Repairer
29N	Telephone Central Office Repairer
29V	Strategic Microwave Systems Repairer
31L	Wire Systems Installer
31V	Unit Level Communications Maintainer
33T	EW/I Tactical Systems Repairer
68J	Aircraft Armament/Missile Systems Repairer
68N	Avionic Mechanic
84B	Still Documentation Specialist
Infantry and Artillery (cluster 5)	
11B	Infantryman
11C	Indirect Fire Infantry man
11H	Heavy Anti-Armor Weapons Infantryman
11M	Fighting Vehicle Infantryman
13B	Cannon Crewmember
13R	FA Firefinder Radar Operator
Military Police (cluster 6)	
95B	Military Police
Operators and Construction (cluster 7)	
16D	Hawk Missile Crewmember
16P	Chaparral Crewmember
44B	Metal Worker
62E	Heavy Construction Equipment Operator
62J	General Construction Equipment Operator
77F	Petroleum Supply Specialist
88M	Motor Transport Operator
Structural Repair and Chemical (cluster 8)	
44E	Machinist
54B	Chemical Operations Specialist

Nine Experimental Job Families (Based on 66 Core Jobs - Only Core Jobs Shown)	
Vehicle and Equipment Repair (cluster 9)	
45K	Tank Turret Repairer
52D	Power Generator Equipment Repairer
62B	Construction Equipment Repairer
63B	Light-Wheel Vehicle Mechanic
63E	M1 Abrams Tank System Mechanic
63G	Fuel and Electrical system Repairer
63S	Heavy-Wheel Vehicle Mechanic
63T	Bradley Fighting Vehicle Systems Mechanic
67V	Observation/Scout Helicopter Repairer
68B	Aircraft Powerplant Repairer
68G	Aircraft Structural Repairer

Mean Prediction Error Scores for 66 MOS

Table 3 shows the mean prediction error (MPE) scores in each of the 66 MOS LSE composites by gender and race. It also shows critical ratios of the differences in PEs for gender and race. At the end of Table 3 are shown the grand (overall) arithmetic and absolute means and overall SDs as well as the number of over- and underpredictions by gender and race.

As noted earlier, the MPE score is equal to the mean AV minus the mean SQT score and the prediction error is computed for each individual. The MPE is computed by groups (i.e., male, female, white and black) within an MOS. The AVs (predicted performance scores) and criterion (SQT) scores were standardized separately within an MOS so as to be on equivalent scales for computation of PEs.

It can be seen at the end of Table 3 that we found underpredictions for blacks in 43 out of the total MOS and underprections for females in 40 out of 50 MOS (females were not assigned to 16 MOS at the time). The overall percentage of underpredictions is around 72 percent for both groups together.

Note, too, at the end of Table 3 we obtained an "adjusted" percentage of the grand absolute means in order to make the larger of female and black scales equivalent in terms of MPE to the male and white scales. This is done by finding the ratio of white to black overall absolute mean and then multiplying the black overall mean that makes it comparable to the white mean. The adjusted value also can be compared across types of AVs within a job family. This procedure is repeated for females.

Table 3

Mean Prediction Error Scores for each of 66 MOS by Gender and Race

	MALE	FEMALE	WHITE	BLACK	CRITICAL RATIO
	Prediction Error Scores				
11B0					
Mean	-0.000	0.000	-0.006	0.029	
SD	0.940	0.000	0.943	0.924	
N	3490	0	2906	584	
%	100.00	0.00	83.27	16.73	
Critical Ratio For Race:					-0.820
Gender:					0.000
11C0					
Mean	-0.000	0.000	0.010	-0.056	
SD	0.934	0.000	0.935	0.926	
N	1896	0	1605	291	
%	100.00	0.00	84.65	15.35	
Critical Ratio For Race:					1.129
Gender:					0.000
11H1					
Mean	-0.000	0.000	-0.010	0.055	
SD	0.937	0.000	0.935	0.947	
N	1027	0	873	154	
%	100.00	0.00	85.00	15.00	
Critical Ratio For Race:					-0.785
Gender:					0.000
11M0					
Mean	0.000	0.000	-0.006	0.026	
SD	0.956	0.000	0.957	0.956	
N	1416	0	1157	259	
%	100.00	0.00	81.71	18.29	
Critical Ratio For Race:					-0.484
Gender:					0.000
12C0					
Mean	0.000	0.000	0.015	-0.070	
SD	0.910	0.000	0.942	0.743	
N	726	0	596	130	
%	100.00	0.00	82.09	17.91	
Critical Ratio For Race:					1.124
Gender:					0.000
13B0					
Mean	-0.000	0.000	0.007	-0.008	
SD	0.903	0.000	0.907	0.899	
N	7851	0	4242	3609	
%	100.00	0.00	54.03	45.97	
Critical Ratio For Race:					0.701
Gender:					0.000

13F0

Mean	-0.000	0.000	-0.017	0.073
SD	0.886	0.000	0.864	0.978
N	1757	0	1431	326
%	100.00	0.00	81.45	18.55

Critical Ratio For Race: -1.535
Gender: 0.000

13M1

Mean	-0.000	0.000	0.024	-0.250
SD	0.943	0.000	0.964	0.653
N	375	0	342	33
%	100.00	0.00	91.20	8.80

Critical Ratio For Race: 2.192
Gender: 0.000

13N1

Mean	0.005	-0.230	0.025	-0.111
SD	0.927	0.713	0.963	0.708
N	463	11	387	87
%	97.68	2.32	81.65	18.35

Critical Ratio For Race: 1.499
Gender: 1.074

13R0

Mean	0.000	0.000	0.048	-0.229
SD	0.972	0.000	1.024	0.633
N	162	0	134	28
%	100.00	0.00	82.72	17.28

Critical Ratio For Race: 1.859
Gender: 0.000

16D0

Mean	-0.016	0.125	0.071	-0.336
SD	0.829	1.105	0.872	0.749
N	247	32	230	49
%	88.53	11.47	82.44	17.56

Critical Ratio For Race: 3.352
Gender: -0.696

16P0

Mean	0.000	0.000	0.003	-0.018
SD	0.916	0.000	0.912	0.946
N	450	0	378	72
%	100.00	0.00	84.00	16.00

Critical Ratio For Race: 0.176
Gender: 0.000

16R2

Mean	-0.000	0.000	-0.003	0.016
SD	0.897	0.000	0.864	1.076
N	399	0	340	59
%	100.00	0.00	85.21	14.79

Critical Ratio For Race: -0.129
Gender: 0.000

16S1

Mean	0.000	0.000	0.019	-0.032	
SD	0.895	0.000	0.859	0.953	
N	837	0	521	316	
%	100.00	0.00	62.25	37.75	
Critical Ratio For Race:					0.785
Gender:					0.000

19E0

Mean	-0.000	0.000	-0.012	0.042	
SD	0.890	0.000	0.888	0.896	
N	1661	0	1298	363	
%	100.00	0.00	78.15	21.85	
Critical Ratio For Race:					-1.008
Gender:					0.000

19K0

Mean	0.000	0.000	-0.004	0.018	
SD	0.874	0.000	0.848	0.985	
N	2714	0	2222	492	
%	100.00	0.00	81.87	18.13	
Critical Ratio For Race:					-0.464
Gender:					0.000

29E0

Mean	0.011	-0.147	0.000	-0.001	
SD	0.849	0.832	0.852	0.827	
N	368	27	331	64	
%	93.16	6.84	83.80	16.20	
Critical Ratio For Race:					0.008
Gender:					0.948

29J0

Mean	0.008	-0.144	0.026	-0.164	
SD	0.897	0.582	0.870	0.963	
N	259	14	235	38	
%	94.87	5.13	86.08	13.92	
Critical Ratio For Race:					1.145
Gender:					0.919

29N0

Mean	-0.009	0.093	-0.052	0.094	
SD	0.894	0.797	0.904	0.847	
N	281	26	198	109	
%	91.53	8.47	64.50	35.50	
Critical Ratio For Race:					-1.410
Gender:					-0.618

29V0

Mean	0.012	-0.111	0.002	-0.018	
SD	0.907	0.810	0.931	0.563	
N	135	14	133	16	
%	90.60	9.40	89.26	10.74	
Critical Ratio For Race:					0.125
Gender:					0.532

31C0

Mean	0.008	-0.081	0.008	-0.040
SD	0.919	0.810	0.902	0.949
N	2587	252	2376	463
%	91.12	8.88	83.69	16.31

Critical Ratio For Race: 1.001
Gender: 1.649

31K0

Mean	0.001	-0.016	0.030	-0.046
SD	0.891	0.888	0.881	0.902
N	2531	219	1667	1083
%	92.04	7.96	60.62	39.38

Critical Ratio For Race: 2.189
Gender: 0.271

31L0

Mean	0.015	-0.057	-0.021	0.023
SD	0.873	0.913	0.837	0.929
N	857	230	570	517
%	78.84	21.16	52.44	47.56

Critical Ratio For Race: -0.804
Gender: 1.068

31V0

Mean	-0.026	0.320	0.046	-0.112
SD	0.902	0.853	0.909	0.880
N	1599	130	1224	505
%	92.48	7.52	70.79	29.21

Critical Ratio For Race: 3.354
Gender: -4.432

33T0

Mean	0.015	-0.349	0.010	-0.665
SD	0.813	0.666	0.808	0.000
N	68	3	70	1
%	95.77	4.23	98.59	1.41

Critical Ratio For Race: 6.987
Gender: 0.918

35K0

Mean	0.026	-0.116	-0.034	0.143
SD	0.938	0.776	0.938	0.778
N	161	36	159	38
%	81.73	18.27	80.71	19.29

Critical Ratio For Race: -1.206
Gender: 0.952

43E0

Mean	0.029	-0.449	0.007	-0.036
SD	0.943	0.962	0.870	1.299
N	354	23	317	60
%	93.90	6.10	84.08	15.92

Critical Ratio For Race: 0.248
Gender: 2.316

```
44B0
Mean    0.002  -0.134   0.021  -0.085
SD      0.806   0.486   0.787   0.857
N         410       7     335      82
%       98.32    1.68   80.34   19.66
Critical Ratio For Race:                  1.017
                 Gender:                  0.726

44E0
Mean    0.003  -0.385   0.006  -0.076
SD      0.800   0.884   0.819   0.488
N         232       2     217      17
%       99.15    0.85   92.74    7.26
Critical Ratio For Race:                  0.627
                 Gender:                  0.618

45K0
Mean    0.013  -0.606   0.003  -0.015
SD      0.831   0.786   0.793   1.026
N         321       7     274      54
%       97.87    2.13   83.54   16.46
Critical Ratio For Race:                  0.121
                 Gender:                  2.060

51B0
Mean   -0.004   0.148   0.007  -0.030
SD      0.897   0.949   0.918   0.808
N         839      20     697     162
%       97.67    2.33   81.14   18.86
Critical Ratio For Race:                  0.519
                 Gender:                 -0.704

52D0
Mean    0.007  -0.153   0.028  -0.101
SD      0.804   0.786   0.803   0.796
N        2285     109    1872     522
%       95.45    4.55   78.20   21.80
Critical Ratio For Race:                  3.255
                 Gender:                  2.084

54B0
Mean    0.016  -0.189  -0.025   0.065
SD      0.815   0.889   0.832   0.794
N         995      83     781     297
%       92.30    7.70   72.45   27.55
Critical Ratio For Race:                 -1.634
                 Gender:                  2.032

55B0
Mean    0.017  -0.176   0.032  -0.084
SD      0.901   0.632   0.900   0.834
N         840      79     664     255
%       91.40    8.60   72.25   27.75
Critical Ratio For Race:                  1.844
                 Gender:                  2.483
```

62B0

Mean	-0.009	0.302	0.007	-0.022
SD	0.736	1.027	0.685	0.927
N	1090	33	864	259
%	97.06	2.94	76.94	23.06

Critical Ratio For Race: 0.457
Gender: -1.727

62E0

Mean	-0.000	0.002	0.006	-0.045
SD	0.868	1.150	0.829	1.128
N	676	7	599	84
%	98.98	1.02	87.70	12.30

Critical Ratio For Race: 0.398
Gender: -0.005

62J0

Mean	0.007	-0.627	0.002	-0.009
SD	0.874	1.172	0.868	0.929
N	378	4	311	71
%	98.95	1.05	81.41	18.59

Critical Ratio For Race: 0.094
Gender: 1.079

63B0

Mean	-0.023	0.230	0.015	-0.041
SD	0.736	0.819	0.722	0.814
N	4040	399	3284	1155
%	91.01	8.99	73.98	26.02

Critical Ratio For Race: 2.064
Gender: -5.926

63E0

Mean	-0.000	0.000	0.023	-0.148
SD	0.870	0.000	0.878	0.806
N	540	0	467	73
%	100.00	0.00	86.48	13.52

Critical Ratio For Race: 1.667
Gender: 0.000

63G0

Mean	0.008	-0.214	0.023	-0.121
SD	0.934	0.794	0.969	0.676
N	300	11	261	50
%	96.46	3.54	83.92	16.08

Critical Ratio For Race: 1.282
Gender: 0.906

63S0

Mean	-0.008	0.463	0.007	-0.087
SD	0.906	1.383	0.912	0.980
N	931	16	873	74
%	98.31	1.69	92.19	7.81

Critical Ratio For Race: 0.800
Gender: -1.358

63T1

Mean	0.000	0.000	0.008	-0.136
SD	0.917	0.000	0.900	1.174
N	700	0	662	38
%	100.00	0.00	94.57	5.43

Critical Ratio For Race: 0.742
Gender: 0.000

67V0

Mean	0.005	-0.235	-0.003	0.039
SD	0.931	1.159	0.921	1.100
N	741	16	696	61
%	97.89	2.11	91.94	8.06

Critical Ratio For Race: -0.293
Gender: 0.821

68B0

Mean	0.001	-0.005	0.017	-0.195
SD	1.040	0.544	1.011	0.414
N	215	41	235	21
%	83.98	16.02	91.80	8.20

Critical Ratio For Race: 1.901
Gender: 0.059

68G0

Mean	-0.008	0.504	-0.002	0.022
SD	0.831	1.366	0.778	1.392
N	378	6	353	31
%	98.44	1.56	91.93	8.07

Critical Ratio For Race: -0.094
Gender: -0.915

68J1

Mean	0.008	-0.230	0.053	-0.271
SD	0.940	0.654	0.960	0.725
N	355	12	307	60
%	96.73	3.27	83.65	16.35

Critical Ratio For Race: 2.992
Gender: 1.220

71D0

Mean	0.114	-0.251	0.014	-0.058
SD	0.915	0.917	0.950	0.846
N	378	172	440	110
%	68.73	31.27	80.00	20.00

Critical Ratio For Race: 0.778
Gender: 4.339

71L0

Mean	0.339	-0.153	-0.026	0.023
SD	0.911	0.888	0.879	0.961
N	238	527	358	407
%	31.11	68.89	46.80	53.20

Critical Ratio For Race: -0.742
Gender: 6.980

71M0					
Mean	0.069	-0.134	-0.017	0.062	
SD	0.884	0.861	0.824	1.072	
N	249	128	298	79	
%	66.05	33.95	79.05	20.95	
Critical Ratio For Race:					-0.608
Gender:					2.144

72E0					
Mean	0.035	-0.131	0.088	-0.123	
SD	0.922	0.845	0.898	0.910	
N	502	136	372	266	
%	78.68	21.32	58.31	41.69	
Critical Ratio For Race:					2.898
Gender:					1.995

72G0					
Mean	0.121	-0.120	-0.009	0.014	
SD	0.973	0.823	0.924	0.885	
N	324	325	388	261	
%	49.92	50.08	59.78	40.22	
Critical Ratio For Race:					-0.316
Gender:					3.403

73C0					
Mean	0.115	-0.148	0.020	-0.024	
SD	0.960	0.887	0.970	0.898	
N	449	350	437	362	
%	56.20	43.80	54.69	45.31	
Critical Ratio For Race:					0.665
Gender:					4.010

74D0					
Mean	0.079	-0.125	-0.006	0.014	
SD	0.886	0.987	0.939	0.911	
N	200	127	236	91	
%	61.16	38.84	72.17	27.83	
Critical Ratio For Race:					-0.175
Gender:					1.896

75B0					
Mean	0.086	-0.184	0.011	-0.012	
SD	0.904	0.885	0.917	0.894	
N	1051	491	821	721	
%	68.16	31.84	53.24	46.76	
Critical Ratio For Race:					0.493
Gender:					5.533

75D0					
Mean	0.090	-0.047	-0.074	0.054	
SD	0.984	0.935	0.913	0.979	
N	337	652	414	575	
%	34.07	65.93	41.86	58.14	
Critical Ratio For Race:					-2.108
Gender:					2.114

76C0					
Mean	0.002	-0.025	0.051	-0.072	
SD	0.889	1.021	0.955	0.803	
N	2263	140	1402	1001	
%	94.17	5.83	58.34	41.66	
Critical Ratio For Race:					3.425
Gender:					0.300
76Y0					
Mean	0.051	-0.266	0.005	-0.007	
SD	0.942	0.839	0.925	0.944	
N	3591	688	2432	1847	
%	83.92	16.08	56.84	43.16	
Critical Ratio For Race:					0.402
Gender:					8.891
77F0					
Mean	-0.003	0.018	-0.025	0.034	
SD	0.806	0.800	0.765	0.854	
N	2456	390	1621	1225	
%	86.30	13.70	56.96	43.04	
Critical Ratio For Race:					-1.909
Gender:					-0.467
81E0					
Mean	0.128	-0.217	-0.029	0.150	
SD	0.843	0.899	0.889	0.815	
N	81	48	108	21	
%	62.79	37.21	83.72	16.28	
Critical Ratio For Race:					-0.908
Gender:					2.156
84B0					
Mean	0.032	-0.244	0.004	-0.028	
SD	0.694	0.964	0.726	0.788	
N	84	11	84	11	
%	88.42	11.58	88.42	11.58	
Critical Ratio For Race:					0.125
Gender:					0.920
84F0					
Mean	0.026	-0.045	0.060	-0.116	
SD	0.926	0.936	0.959	0.858	
N	58	33	60	31	
%	63.74	36.26	65.93	34.07	
Critical Ratio For Race:					0.886
Gender:					0.350
88H0					
Mean	0.022	-0.163	0.081	-0.118	
SD	0.954	0.864	0.932	0.953	
N	469	64	316	217	
%	87.99	12.01	59.29	40.71	
Critical Ratio For Race:					2.386
Gender:					1.585

88M0					
Mean	0.007	-0.055	0.011	-0.025	
SD	0.843	0.832	0.827	0.873	
N	4758	610	3694	1674	
%	88.64	11.36	68.82	31.18	
Critical Ratio For Race:					1.422
Gender:					1.720

91A0					
Mean	0.046	-0.230	-0.013	0.036	
SD	0.904	0.936	0.910	0.930	
N	1493	297	1306	484	
%	83.41	16.59	72.96	27.04	
Critical Ratio For Race:					-1.003
Gender:					4.674

94B0					
Mean	0.052	-0.222	0.045	-0.040	
SD	0.905	0.876	0.984	0.830	
N	3069	718	1772	2015	
%	81.04	18.96	46.79	53.21	
Critical Ratio For Race:					2.858
Gender:					7.483

95B0					
Mean	0.027	-0.179	-0.009	0.099	
SD	0.921	0.877	0.912	0.977	
N	2059	310	2182	187	
%	86.91	13.09	92.11	7.89	
Critical Ratio For Race:					-1.454
Gender:					3.820

	MALE	FEMALE[1]	WHITE	BLACK[2]
GRAND ARITHMETIC MEAN	0.012	-0.108	0.008	-0.019
GRAND ABSOLUTE MEAN	0.016	0.150	0.017	0.040
GRAND SD	0.060	0.409	0.068	0.192
PREDICTIONS				
OVER	40	10	43	22
UNDER	9	40	22	43

[1] Adjusted Mean for Female: -0.012
[2] Adjusted Mean for Blacks: -0.008

Test Composites With Statistically Significant MPEs For Blacks

Table 4 shows the thirteen MOS LSE with one over- and 12 underpredictions for blacks that have a statistically significant critical ratio at the .05 level. This is not strictly a test of statistical significance because each MOS was tested individually rather than as one in a set of 66 MOS with highly correlated AVs. In such a set, five or more ratios significant at the .05 level could be expected by chance (McNemar, 1949).

The table shows that the mean prediction error for each of the thirteen MOS ranges from –0.665 to .054. The overall mean prediction error across the 66 MOS is only –0.019. As a frame of reference for these values, which were computed using a standard score scale having a standard score mean of zero and a SD of one, we can convert these values to an aptitude area composite score having a mean of 100 and a SD of 20. The prediction correction column of Table 4 gives the correction required to make the error equal to zero for the black group. For example, the –.019 (overall mean across the 66 MOS) is converted to a Army standard score of 0.38, or less than 1 point on an Army standard score scale, to make the mean prediction error (AVs –SQTs) equal zero for the total set of MOS.

Examining the thirteen MOS with significant critical ratios, the range of corrections in AA standard score units required varies between –6.72 and 1.08 (excluding 33T with one observation). We consider this magnitude of correction in a classification context of little practical consequence.

Critical ratios for each MOS were transformed into chi squares and then tested for statistical significance as a set (McNemar, 1949). This was done by summing all MOS with negative mean prediction error signs (43 MOS). MOS scores with positive signs were assumed to be zero, which is a very conservative assumption for this test. The overall chi square of the set

was found to be statistically significant at the .05 level. However, the findings for the composites, taken as a whole, show that the set of mean prediction error scores, while showing statistically significant chi squares at the .05 level for blacks, were of little or no practical significance. Further, there was a lack of any pattern or association among the 13 MOS pointing to experimental error.

Table 4

Statistically Significant Mean Prediction Errors for Blacks

		N	Prediction Error	Prediction Correction
	Over-prediction for Blacks			
75D0	Personnel Records Specialist	575	0.054	1.08
	Under-predictions for Blacks			
13M1	Multiple Launch Rocket Sys (MLRS) Crewmember	33	-0.25	-5
16D0	Hawk Missile Crewmember	49	-0.336	-6.72
31K0	Combat Signaler	1083	-0.046	-0.92
31V0	Unit Level Communications Maintainer	505	-0.112	-2.24
33T0	EW/I Tactical Systems Repairer	1	-0.665	-13.3
52D0	Power Generator Equipment Repairer	522	-0.101	-2.02
63B0	Light-Wheel Vehicle Mechanic	1155	-0.041	-0.82
68J1	Aircraft Armament/Missile Systems Repairer	60	-0.271	-5.42
72E0	Tactical Telecommunications Center Op	266	-0.123	-2.46
76C0	Equipment Records and parts Specialist	1001	-0.072	-1.44
88H0	Cargo Specialist	217	-0.118	-2.36
94B0	Food Service Specialist	2015	-0.04	-0.80
Total for 66 MOS		24698	-0.019	0.38

Statistically Significant Test Composites for Females

Table 5 presents the same type of data for females as does the previous table for blacks. In this table, we find two MOS that were overpredicted and 17 that were underpredicted. Here, the overall mean prediction error is –0.108. The 19 MPEs have statistically significant critical ratios at the .05 level. Again, this is not strictly a test of statistical significance because each MOS was tested individually rather than as one set of 50 MOS. In such a set, five or more ratios significant at the .05 level could be expected by chance (McNemar, 1949).

The overall chi square for gender for a set of 50 MOS is statistically significant ($p < .05$), assuming independence. However, again, it is not of practical significance. The mean predicted error is –0.108 or 2.16 Army standard score units (Table 5). It is worth noting that 47 percent of the total female sample is black, which may be confounding gender and race results. Since prediction error scores are more negative for females than for blacks in the MOS shown in Table 5, gender effects seem to be dominant.

Examining the MOS content of over and underpredictions for females in Table 5, we find a pattern. The overpredictions are for two technical MOS that provide substantial weight to technical information tests in the AV composites. Except for the Food Service and several other MOS, most of the other MOS showing significant underpredictions fall within the Administrative / Clerical job families.

Table 5

Statistically Significant Mean Prediction Errors for Females

		N	Prediction Error	Prediction Correction
	Over-prediction for Females			
31V0	Unit Level Communications Maintainer	130	0.32	6.4
63B0	Light-Wheel Vehicle Mechanic	399	0.23	4.6
	Under-predictions for Females			
43E0	Parachute Rigger	23	-0.449	-8.98
45K0	Tank Turret Repairer	7	-0.606	-12.12
52D0	Power Generator Equipment Repairer	109	-0.153	-3.06
54B0	Chemical Operations Specialist	83	-0.189	-3.78
55B0	Ammunitions Specialist	79	-0.176	-3.52
71D0	Legal Specialist	172	-0.251	-5.02
71L0	Administrative Specialist	527	-0.153	-3.06
71M0	Chaplain Assistant	128	-0.134	-2.68
72G0	Automatic Data Telecommunications Center Op	325	-0.12	-2.4
73C0	Finance Specialist	350	-0.148	-2.96
75B0	Personnel Administration Specialist	491	-0.184	-3.68
75D0	Personnel Records Specialist	652	-0.047	-0.94
76Y0	Unit Supply Specialist	688	-0.266	-5.32
81E0	Graphics Documentation Specialist	48	-0.217	-4.34
91A0	Medical Specialist	297	-0.23	-4.6
94B0	Food Service Specialist	718	-0.222	-4.44
95B0	Military Police	310	-0.179	-3.58
Total for 66 MOS		8088	-0.108	2.16

Mean Prediction Error Scores for Nine Experimental Job Families

Table 6 shows the mean prediction error scores (MPEs) for each of the composites of seven tests by gender and race for the nine experimental families. The MPE for blacks is –0.004, and for females is –0.036.

We found four statistically significant overpredictions and three statistically significant underpredictions for blacks (see bottom of Table 6) across the nine job families. For females there were three significant over- and five significant underpredictions. There were two families that were neither over- nor underpredicted for blacks and one such family for females. The largest overprediction error for blacks was .073 and for females was .126. The largest (absolute value) underprediction for blacks was –0.049 and for females was –0.264.

The grand PE standard deviation for males was small, .038, as compared to females at .291. For the 66 MOS data set shown in Table 3, we again find much less variability for blacks than for females with a grand standard deviation for PEs of .192 versus .409, respectively.

Tables 7 and 8 show the specific families from among the 9 experimental job families that had statistically significant critical ratio at the .05 level, following the same procedure described for Tables 3 and 4. The largest prediction correction to make the PE equivalent to zero is about .08 points for blacks and .72 for females on an AA scale with a mean of 100 and SD of 20. The chi-squares of the two sets proved to be not significant for both blacks and females. We consider the magnitude of PEs in Tables 7 and 8 to be of little practical consequence.

Table 6

Mean Prediction Error Scores for each of the Nine Experimental Job Families by Gender and Race

	Prediction Error scores				
	MALE	FEMALE	WHITE	BLACK	CRITICAL RATIO
Administration (cluster 1)					
Mean	0.016	-0.044	-0.002	0.002	
SD	0.928	0.955	0.931	0.942	
N	10527	3765	8407	5885	
%	73.66	26.34	58.82	41.18	
Critical Ratio For Race:					-0.243
Gender:					3.299
Armor and Combat Specialties (cluster 2)					
Mean	-0.009	0.126	-0.028	0.073	
SD	0.883	0.854	0.886	0.866	
N	9627	652	7461	2818	
%	93.66	6.34	72.58	27.42	
Critical Ratio For Race:					-5.207
Gender:					-3.887
Combat Systems Operations (cluster 3)					
Mean	0.006	-0.057	0.017	-0.038	
SD	0.911	0.698	0.919	0:834	
N	10681	1083	8150	3614	
%	90.79	9.21	69.28	30.72	
Critical Ratio For Race:					3.196
Gender:					2.736
Electronics Repair (cluster 4)					
Mean	-0.007	0.061	0.000	-0.001	
SD	0.893	0.837	0.905	0.844	
N	4167	503	3311	1359	
%	89.23	10.77	70.90	29.10	
Critical Ratio For Race:					0.042
Gender:					-1.720
Infantry and Artillery (cluster 5)					
Mean	-0.000	0.000	-0.005	0.012	
SD	0.923	0.000	0.932	0.901	
N	15842	0	10917	4925	
%	100.00	0.00	68.91	31.09	
Critical Ratio For Race:					-1.086
Gender:					0.000
Military Police (cluster 6)					
Mean	0.027	-0.179	-0.009	0.099	
SD	0.921	0.877	0.912	0.977	
N	2059	310	2182	187	
%	86.91	13.09	92.11	7.89	
Critical Ratio For Race:					-1.454
Gender:					3.820

Operators and Construction (cluster 7)

Mean	0.014	-0.125	0.022	-0.049
SD	0.864	0.811	0.853	0.873
N	9375	1050	7168	3257
%	89.93	10.07	68.76	31.24

Critical Ratio For Race: 3.929
Gender: 5.243

Structural Repair and Chemical (cluster 8)

Mean	0.018	-0.264	0.000	-0.000
SD	0.833	0.867	0.857	0.774
N	1227	85	998	314
%	93.52	6.48	76.07	23.93

Critical Ratio For Race: 0.013
Gender: 2.907

Vehicle and Equipment Repair (cluster 9)

Mean	-0.003	0.050	0.009	-0.039
SD	0.828	0.890	0.833	0.822
N	11541	638	9841	2338
%	94.76	5.24	80.80	19.20

Critical Ratio For Race: 2.559
Gender: -1.458

	MALE	FEMALE[1]	WHITE	BLACK[2]
GRAND ARITHMETIC MEAN	0.004	-0.036	0.002	-0.004
GRAND ABSOLUTE MEAN	0.008	0.072	0.012	0.028
GRAND SD	0.038	0.291	0.037	0.062
PREDICTIONS OVER	5	3	3	4
PREDICTIONS UNDER	3	5	4	3

[1] Adjusted Mean for Female: -0.004
[2] Adjusted Mean for Blacks: -0.002

Statistically Significant Critical Ratios for Nine Experimental Families (AAs) for Blacks and Females

Table 7

Statistically Significant Mean Prediction Errors for Blacks

	N	Prediction Error	Prediction Correction
Cluster	Over-prediction for Blacks		
2	2818	0.073	1.46
	Under-predictions for Blacks		
3	3614	-0.038	-0.76
7	3257	-0.049	-0.98
9	2338	-0.039	-0.78
Total for 9 AA	24698	-0.004	0.08

Table 8

Statistically Significant Mean Prediction Errors for Females

	N	Prediction Error	Prediction Correction
Cluster	Over-prediction for Females		
2	652	0.126	2.52
	Under-predictions for Females		
1	3765	-0.044	-0.88
3	1083	-0.057	-1.14
6	310	-0.179	-3.58
7	1050	-0.125	-2.5
8	85	-0.264	-5.28
Total for 9 AA	8088	-0.036	0.72

Comparison of Prediction Error Scores

1. In Table 3 of the present study, the grand arithmetic mean for the 66 MOS seven test composites was –0.019 for blacks. For females, the value was –0.108. For the previous study (Zeidner et al., 1998), the values were –0.025 for blacks and –0.086 for females (Table 6). We consider the results from all of these comparisons of little practical importance.

2. In Table 6 of the present study, the grand arithmetic mean for the nine experimental job families (composites of seven tests) was -.004 for blacks. For females, the PE value was -.036. For the previous study the PE values were -.012 for blacks and -.006 for females (Table 12).

3. In Table 3 of the present study, for the 66 MOS, examining composites using seven tests, the grand SD for PEs was .192 for blacks and .409 for females. In the previous study, for the 66 MOS test composites the grand PE SD was .310 for blacks and .440 for females (Table 6). Thus for both sets of comparisons, PE SDs for females were considerably more variable.

4. There were underpredictions in 43 of 66 MOS (65 percent) for blacks in the present study (Table 3). For females there were statistically significant underpredictions in 40 of 50 MOS (80 percent). Similar findings were found for the previous study.

5. There were significant underpredictions of three of nine (33 percent) MOS for blacks in the present study of nine job families, seven test composites (Table 6). For females, five of eight (or 63 percent) were significant underpredictions (there were no females in cluster 5). Again, there are similar results for the composites based on nine tests of the previous study.

6. In sum, for the six comparisons of this study across 66 MOS and nine experimental job families: (1) females were more underpredicted, (2) females had more variability in PEs than blacks, and (3) in the critical area of the size of PEs, females and blacks were considered to be comparable in terms of practical significance. For example, the overall MPE for the 13 (of 66) statistically significant MOS for

blacks is -.019 (Table 4), or less than one half of an Army standard score point in the AA system (mean of 100 and a SD of 20). Females have higher statistically significant MPEs, -0.108, and have more MOS that are underpredicted (Table 5). About two standard score points in the Army standard score scale would correct the MPE value to zero. In terms of MPEs for blacks and females, the differences in the size of PEs is of little practical importance.

Conclusions

Although we found the size of mean prediction errors to be of little or no practical significance as in the previous study, we were puzzled by the consistent trend of underpredictions for blacks and females in contrast to the consistent overpredictions cited by the literature. Our findings of lower predicted performance values were consistent with many, if not most, findings citing lower intercept values for blacks and females. But most of these same studies suggested that minorities were overpredicted when total sample data (or regression lines) were used. Our total sample data, in contrast, showed underpredictions. We offer two possible explanations of the disparity of findings, one of which is based on psychometric analyses, the other on a motivational hypothesis.

Many studies, especially in the civilian employment sector, reporting overpredictions for minorities when total sample data are used appear to have a relatively small percentage of minorities in the total sample. Consequently, predicted performance scores of minorities in such studies may be increased relative to the present study where minorities (females and blacks) constitute about 40 percent of the total sample. The relatively smaller increases in predicted performance scores of minorities scaled in terms of the total Army input sample, as in the present study, were found not to be large enough to result in overpredictions.

A second explanation may also partially contribute to underpredictions for minorities. Army careers are greatly valued by many minority Soldiers. The Army has a much higher percentage of blacks than do other services and is generally considered to have good promotional opportunities for minorities. For

example, a higher percentage of blacks than whites reenlist in the Army. We hypothesize, then, that blacks, in general, are very motivated to perform well in their first terms of duty and desire to remain in the career force. Blacks also are provided with the same type of formal training as whites and appear to have similar on-the-job learning experiences as whites. Motivation and equal opportunity to learn for minorities may contribute to higher actual performance, as measured by SQT, than scores predicted by the cognitive ability test composites of ASVAB. The same motivational hypothesis is offered for females, nearly half of whom are black. But again, we note that the prediction error scores of minorities are very small and of little or no practical significance.

Overall, then, we conclude that the composites of seven weighted tests can be considered fair for predicting job performance with ASVAB for both gender and race in the context of today's operational selection system even though such test composites provide less accuracy than were provided by composites of nine tests.

We also conclude that the LSE composites using seven tests can be considered fair in the classification context and that they provide for greater accuracy than did the former unit-weighted AAs. However, the loss of the CS and NO tests seems to have more of a negative effect on females in terms of the increased number of MOS that have statistically significant MPEs. Again, for practical purposes, the magnitude of MPEs is small and quite comparable for females and blacks.

REFERENCES

Cleary, T. A. (1968). Test bias: Prediction of grades for Negro and white students in integrated colleges. *Journal of Educational Measurement, 5,* 115-124.

McNemar, Q. (1949). *Psychological Statistics.* New York: John Wiley & Sons.

Thorndike, R.L. (1971). Concepts of Cultural Fairness. *Journal of Educational Measurement, 8, 63-70.*

Wise, L., Welsh, J., Grafton, F., Foley, P., Earles, J., Sawin, L., & Divgi, D. R. (1992). *Sensitivity and Fairness of the Armed Services Vocational Aptitude Battery (ASVAB) Technical Composites* (DMDC Technical Report 92-002). Monterey, CA: Defense Manpower Data Center.

Zeidner, J., Johnson, C.D., Vladimirsky, Y., & Weldon, S. (August 2000). *Specifications for an Operational Two-Tiered Classification System for the Army, Volume 1: Report* (Technical Report 1108). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Zeidner, J., Johnson, C.D., Vladimirsky, Y., & Weldon, S. (December 2000). *The Effect of Reducing the Number of Tests in the Armed Services Vocational Aptitude Battery* (Study Note 2001-01). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Zeidner, J., Johnson, C.D., Vladimirsky, Y., & Weldon, S. (November 2002). *Determining Composite Validity Coefficients for Army Jobs and Job Families* (Study Note 2003-02). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Zeidner, J., Johnson, C.D., Vladimirsky, Y., & Weldon, S. (January 2003). *Determining Mean Predicted Performance for Army Job Families* (Study Note 2003-03). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Zeidner, J., Johnson, C.D., Vladimirsky, Y., & Weldon, S. (1998). *Fairness of Proposed New Army ASVAB Test Composites for Restructured Job Families.* Contractor report submitted to ARI.

Zeidner, J., Scholarios, D., & Johnson, C. (2003). *Evaluating Job Knowledge Criterion Components for use in Classification Research. Military Psychology.* Erlbaum Associates. Mahway, NJ.